

PROJET MICRO : MODELISATION INFORMATIQUE DE LA COGNITION EN RECONNAISSANCE DE L'ORAL

Jean CAELEN, Bertrand CAILLAUD, Jean-Yves ANTOINE

Institut de la Communication Parlée
URA CNRS 368, INPG
46, Avenue Felix Viallet, 38 031 Grenoble Cedex
Email: caillaud@icp.grenet.fr, antoine @icp.grenet.fr, jcaelen@icp.grenet.fr

Résumé

*

Cet article propose une architecture de système de compréhension automatique de la parole inspirée de modèles cognitifs. Les résultats acquis en neuropsychologie et en psycholinguistique nous ont poussé à doter notre système d'une architecture distribuée constituée d'un réseau d'agents. De plus, compte tenu de la latéralisation des processus cérébraux dans les activités langagières, nous avons choisi d'ajouter au traitement analytique de la parole une seconde voie rapide reposant sur l'analyse prosodique du signal et dont la finalité est de fournir des points d'ancrages temporels à ces processus. Après avoir détaillé les théories cognitives ayant inspiré notre recherche, nous décrivons l'architecture globale du système, en nous attachant sur la description de chaque niveau d'analyse. Enfin, nous présentons une étude de stratégie de contrôle effectuée à l'aide d'une simulation du système.

Mots-clés : système cognitif, compréhension de la parole, Intelligence Artificielle Répartie, système multi-agents, analyse holistique.

Abstract

This paper focuses on the description of the MICRO system, a speech understanding system largely inspired by cognitive models. The aim of this project is to improve the adaptive abilities of the system by means of this cognitive modelization. Following cognitive modularist theories, MICRO was thus given a multi-agent structure. Furthermore, we took into account the functional difference between analytic and holistic processing that has been characterized in human cognition. Consequently, a fast (prosodic) way of analysis, devoted to the definition of anchorage points, was added to the classical analytic strategy (from acoustic-phonetic decoding to linguistic analysis). We first present a detailed overview of cognitive theories that represent the background of the project. The structure of MICRO system is then widely described. We finally describe the realization of the simulation of MICRO, aimed at studying several strategies of control.

Keywords : cognition, speech understanding Distributed Artificial Intelligence, multi-agents system, holistic analysis.

¹* Modélisation Informatique de la Cognition en Reconnaissance de l'Oral

1. PRESENTATION

1.1. GENESE DU PROJET

Dans le domaine de la reconnaissance de la parole, les systèmes fondés sur des techniques de reconnaissance structurelle des formes (modèles de Markov par exemple) ont acquis ces dernières années un *leadership* en matière de performances d'ensemble que ne peuvent leur contester les systèmes d'Intelligence Artificielle fondés sur les connaissances. Néanmoins, les systèmes auto-organiseurs ou statistiques actuels sont incapables d'aborder correctement les étapes de compréhension des énoncés pour deux raisons majeures : (a) ils souffrent d'une capacité de représentation limitée [19] préjudiciable à la réalisation de processus d'interprétation robustes, et (b) la communication orale se caractérisant par une occurrence fréquente de structures agrammaticales et irrégulières, ils souffrent également d'un pouvoir de généralisation limité. La prise en compte de ces énoncés non normés nécessite un comportement de recherche hautement adaptatif, capacité dont sont dépourvus les systèmes de reconnaissance structurelle. Or, l'être humain présente de remarquables capacités d'adaptation et d'apprentissage par généralisation : il raisonne par abduction ou par analogie. On part de l'hypothèse que la faiblesse des systèmes actuels provient de leur absence de modélisation de ces niveaux de raisonnement.

1.2. CADRE DE RECHERCHE

Le domaine d'application choisi pour le système de compréhension automatique de la parole MICRO est la conception de plans architecturaux assistée par ordinateur (ICPplan). Cette tâche particulière de conception engendre l'utilisation d'une grande variété de structures linguistiques du fait que le concepteur formule ses commandes dans le cours de l'action. On note en particulier des hésitations et des reprises nombreuses. Le dialogue d'ICPplan est multimodal, des anaphores et des déictiques apparaissent fréquemment. Nous visons dans ce projet la compréhension d'énoncés en parole naturelle sur un lexique d'environ 200 mots en multilocuteur.

2. APPROCHE COGNITIVE

Après une étude bibliographique, nous avons retenu trois caractéristiques essentielles des systèmes cognitifs utiles à notre propos : la modularité, l'interactivité, les deux voies d'analyse.

2.1. MODULARITE

Le paradigme modulariste constitue le cadre principal de notre modélisation. Il présente la cognition comme la résultante de l'activité coopérative d'un ensemble de modules *indépendants* travaillant des domaines de compétence propres. La détermination de plus en plus précise de localisations fonctionnelles du cerveau humain par les neurosciences et la psychologie expérimentale est à la base de la thèse modulariste. On distingue par exemple dans le cortex visuel primaire des aires spécifiques effectuant en parallèle une ségrégation fonctionnelle de l'information rétinienne (couleur, direction...) [18]. Cette approche répartie se retrouve, en moins marquée, en perception, mais il n'est guère possible de localiser physiologiquement des fonctions cognitives supérieures, du fait de la plasticité du système nerveux. La psychologie expérimentale a néanmoins caractérisé fonctionnellement certains modules cognitifs supérieurs. Marr propose ainsi un modèle modulaire de vision [23], de même que Kosslyn [21] qui distingue plusieurs modules travaillant en parallèle suivant deux voies de traitements physiologiquement différenciées. Enfin, Morton propose un modèle réparti des activités langagières [29]. Ces travaux ont ainsi conduit à l'émergence de plusieurs théories modularistes générales de la cognition. Fodor, s'est ainsi - timidement - rallié à cette thèse et décrit un ensemble de modules *périphériques* (bas-niveau) construisant des représentations utilisées par des systèmes centraux non modulaires [12]. De même, Minsky conçoit désormais la cognition comme la résultante de l'activité de myriades d'éléments non conscients organisés en société [27].

L'Intelligence Artificielle a, dans un premier temps, modélisé ce paradigme à l'aide architectures à tableau noir, composées d'une mémoire partagée permettant une communication entre divers experts [17]. Cette architecture imposait néanmoins un contrôle centralisé préjudiciable à l'émergence d'un comportement adaptatif. L'Intelligence Artificielle Répartie a alors proposé de distribuer le contrôle entre les agents du système, comme dans le système MICRO, qui se présente comme un système multi-agents hétérarchique [1],[10].

2.2. INTERACTIVITE

Les Sciences Cognitives n'ont pas tranché le vieux débat *empiriste/constructivite* à propos de la stratégie de résolution suivie par le système cognitif : (a) bottom-up (construction ascendante de la connaissance à partir de l'information perceptive), ou (b) top-down (élaboration des connaissances sous le contrôle de niveaux supérieurs), ou (c) mixte ?

Les autonomistes, tel Fodor, plaident pour une stratégie ascendante résultant de l'encapsulation des modules cognitifs qui empêche tout retour des systèmes centraux vers les modules inférieurs. Ce paradigme présente surtout une justification expérimentale au niveau perceptif [36]. A l'opposé, la théorie interactionniste [25] insiste sur la prise en compte simultanée des informations ascendantes et descendantes. On sait ainsi qu'il existe des connexions de retour entre aires visuelles supérieures et inférieures [18]. Parallèlement, les activités langagières impliquent un comportement intégré dans lequel il est bien difficile de distinguer stratégies ascendante et descendante [1]. De nombreux modèles présentent ainsi l'accès lexical comme un processus interactif [26],[24]. L'interactivité présentant un intérêt certain d'un point de vue adaptatif (cf § 2.4.), nous avons doté le système MICRO d'une capacité de traitement interactif réalisée par des processus diffusants ascendants et/ou descendants indifférenciés.

2.3. HEMISPHERE HOLISTIQUE ET HEMISPHERE ANALYTIQUE

En affirmant que la perception dépendait de l'organisation globale des stimuli, et non de la combinaison de traits perceptifs, la psychologie gestaltiste a posé très tôt la question *holistique*. Gardner propose une définition analogue lorsqu'il distingue processus perceptifs analytiques, traitant les dimensions physiques séparables et processus holistiques, analysant les dimensions intégrables formant un tout [14]. Les processus holistiques effectuent ainsi une analyse globale et synthétique des représentations, à la différence des processus analytiques.

Sperry a pu montrer une différenciation fonctionnelle entre l'hémisphère gauche langagier et l'hémisphère droit non verbal [35]. La réalité est peut-être moins tranchée [6]. Quoiqu'il en soit cette spécialisation hémisphérique résulterait de la nature propre de chaque analyse hémisphérique [15] : "l'hémisphère droit serait spécialisé dans un style de pensée synthétique de type Gestalt, alors que l'hémisphère gauche suivrait des stratégies de pensée fragmentaires, analytiques" [37]. L'hémisphère gauche a ainsi été vu comme centre du langage parce que les traitements phonétiques et linguistiques sont dominés par des processus analytiques. Au contraire, divers travaux ont montré que l'analyse de la prosodie est dominée par l'hémisphère droit [39], principalement holistique. Nous avons alors défini dans MICRO une voie d'analyse prosodique effectuant un pré-marquage rapide d'événements pour une deuxième voie d'analyse plus classique.

2.4. MODELISATION COGNITIVE ET ADAPTATION

D'un point de vue adaptatif, c'est surtout l'indépendance des processus cognitifs qui nous semble primordiale dans le paradigme modulariste. Cette indépendance permet en effet de rapprocher les pôles de décisions des sources d'adaptations contextuelles. De même, l'interactivité autorise une prise en compte précoce des informations contextuelles fournies par les modules supérieurs. Cet aspect est de première importance en dialogue, où la caractérisation de l'intention du locuteur permet d'effectuer des prédictions linguistiques. Néanmoins, nous insisterons surtout sur la coopération entre analyses différentes (holistique et analytique dans MICRO) qui permet un comportement auto-organisateur [22]. Lautrey distingue en effet trois types de relations entre processus cognitifs:

- *vicariance* : les processus remplissent une fonction analogue à partir de la même information.

- *complémentarité* : les processus vicariants travaillent à partir d'informations différentes.
- *interaction* : les processus vicariants ou complémentaires sont tels que l'action de l'un modifie celle des autres par la médiation d'une représentation partagée.

Selon Lautrey, les relations de vicariance sont invoquées pour l'adaptation dans des situations habituelles, les situations rares mettant en jeu la relation de complémentarité. Enfin, les situations nouvelles sont gérées par les relations d'interaction. On retrouve cette distinction entre un hémisphère droit gérant les situations usuelles alors que l'hémisphère gauche prend en compte les situations nouvelles [37]. Cette complémentarité holistique-analytique est esquissée par MICRO.

3. DESCRIPTION DE MICRO

Le système MICRO est développé dans l'environnement MAPS, une plateforme générique dédiée à la conception de systèmes experts de type multi-agents [2]. L'architecture de MICRO ne peut se comprendre sans une description sommaire de cet environnement.

3.1. MAPS

La conception de MAPS repose sur la distinction entre deux classes de connaissances :

- *les connaissances descriptives* sur les éléments du problème (données, hypothèses, résultats),
- *les connaissances opératoires et décisionnelles* (règles et meta-règles), définissant des opérations et un mécanisme de raisonnement sur les éléments du problème.

MAPS associe un type d'agent à chaque classe de connaissance, respectivement les KS (Knowledge Server) et les KP (Knowledge Processor). Les agents communiquent par envoi de messages et réagissent ainsi à des événements extérieurs. Les agents KS communiquent usuellement avec des agents KP (soumission de données et problèmes), et réciproquement les agents KP avec des agents KS (transmission et demande d'informations). MAPS permet une exécution parallèle des processus attachés aux agents ; deux modes de synchronisation (synchrone et asynchrone) entre les agents ont été introduits pour permettre la définition de différents comportements (agents indépendants, coordonnés ou concurrents). Les communications en mode synchrone correspondent à des relations *client / serveur*, tandis qu'en mode asynchrone, les communications sont de type *boîte aux lettres*. Enfin, MAPS permet désormais la création d'architectures multi-couches par l'introduction de méta-agents, chargés du contrôle respectif des tâches et des connaissances.

3.2 ARCHITECTURE GLOBALE DU SYSTEME MICRO

Le système est décrit en figure 1 (voir [10] pour une description détaillée). On remarque la structure hétérarchique du système qui est fractionné en plusieurs groupes d'agents correspondant à des niveaux de traitement spécifiques. On distingue les groupes phonétique et linguistique pour la voie analytique, et enfin un groupe prosodique comprenant l'ensemble des agents de la voie de pré-marquage holistique. L'analyse acoustique est effectuée à l'aide d'un modèle mathématique d'oreille [7] comportant plusieurs étages de traitement.

Nous détaillons seulement dans les paragraphes suivants la structure des autres groupes d'agents.

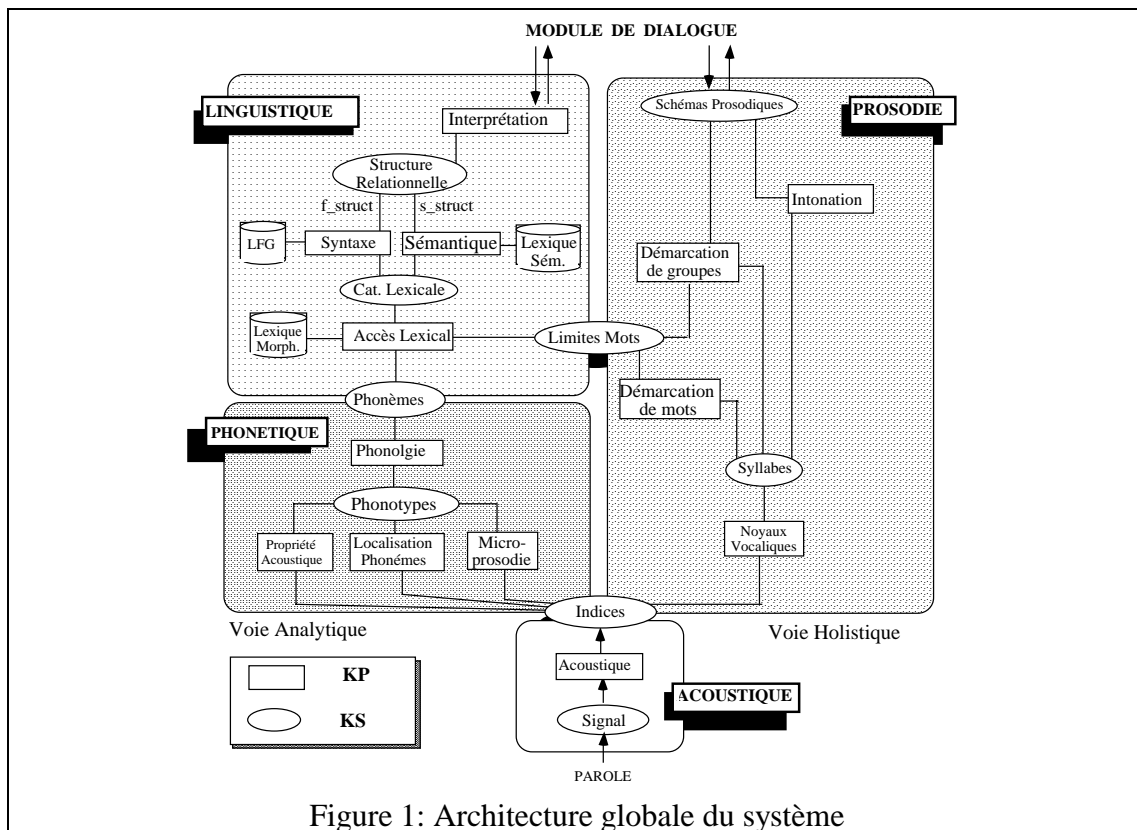


Figure 1: Architecture globale du système

4. GROUPE ACOUSTICO-PHONETIQUE

Ce groupe d'agents rassemble les traitements bas-niveau de la voie analytique. Les agents qui le composent doivent produire des hypothèses phonétiques, à partir d'une représentation en traits de la parole (fournie par le groupe acoustique) et des contraintes phonétiques provenant du groupe linguistique. L'expérience acquise avec le système DIRA [8] a permis une détermination aisée de la structure de ce groupe (figure 2). Il est constitué de cinq agents internes et de deux KS (en grisé sur la figure) qui servent de relais avec les deux autres groupes voisins (KS Phonème et Indices).

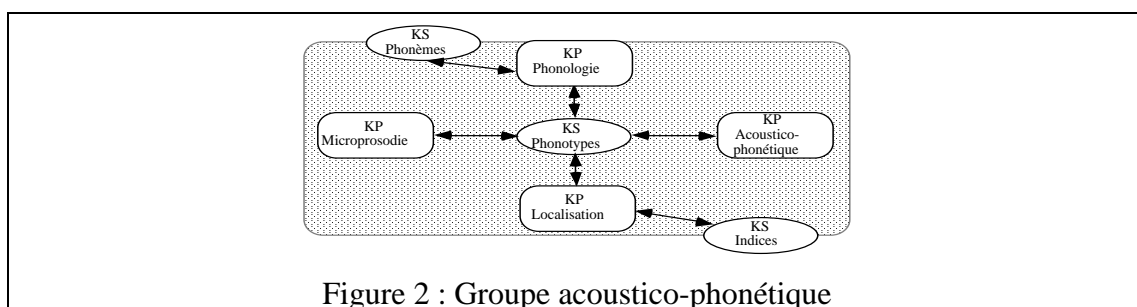


Figure 2 : Groupe acoustico-phonétique

4.1. KP LOCALISATION

Cet agent détermine les localisations temporelles des phonotypes² à partir d'une sur-segmentation du signal effectuée par le groupe acoustique. Les segments obtenus sont de type *phones homogènes*. Les agents du groupe acoustico-phonétique travaillent tous sur ces segments.

4.2. KP ACOUSTICO-PHONETIQUE

Cet agent classe les phonotypes à l'aide des traits acoustico-phonétiques suivants :

- *Macro-classe* : trait de catégorisation acoustico-phonétique large. Les macro-classes sont : occlusif, fricatif, vocalique, consonantique, silence.

²phonotype: c'est la réalisation acoustique d'un phonème.

- *Traits*: (aigu-grave, fermé-ouvert, doux-strident, bémolisé-diésé, écarté-compact) : Ils apportent des informations acoustiques complémentaires aux macro-classes, permettant une restriction éventuelle de l'espace de recherche.

4.3. KP MICROPROSODIE

Cet agent caractérise les phonotypes à l'aide des traits micro-prosodiques suivants :

- *Voisement* : trait de catégorisation binaire (voisé / non voisé) sur le voisement du phonotype.
- *Energie* : trait de catégorisation micro-prosodique décrivant la forme de la courbe d'énergie. Il est analogue, sur son domaine de compétence, au trait "Macro-classe".
- *Durée intrinsèque* : trait qualifiant la durée du phonotype par rapport à une durée étalon.

4.4. KP PHONOLOGIE

Cet agent a une production descendante vis à vis du groupe, contrairement aux autres agents. Il produit des hypothèses phonotypiques à partir des contraintes produites par le groupe linguistique.

4.5. KS PHONOTYPES :

Cet agent mémorise les informations produites par les KP du groupe.

5. NIVEAU LINGUISTIQUE

5.1. LINGUISTIQUE ET DIALOGUE

Les systèmes stochastiques présentent une bonne robustesse en reconnaissance phonétique et lexicale. A l'opposé, l'analyse linguistique ne peut s'envisager correctement qu'à l'aide de systèmes à bases de connaissances. De plus, le caractère formel de l'analyse syntaxique se prête aisément au formalisme des systèmes à base de règles. Néanmoins, le schéma formel classiquement envisagé en TALN (Traitement des Langues Naturelles) ne peut s'appliquer au domaine du dialogue oral. En effet, la forte proportion de phrases grammaticalement incorrectes (hésitations, reprises, élisions) rend inappropriée toute analyse syntaxique. Cette spécificité de la *linguistique dialogique* nécessite la prise en compte du sujet parlant et de la situation. Sans entrer totalement dans la pragmatique, nous avons conçu une sémantique finalisée, exposée ci-après.

5.2. COOPERATION SYNTAXE-SEMANTIQUE

Il est remarquable que l'agrammaticalité des énoncés de dialogue n'empêche pas leur compréhension. Rattcliff (cité dans [13]) a ainsi montré que les sujets peuvent effectuer des décisions de sens sur des phrases agrammaticales. L'intervention des processus sémantiques au cours de l'analyse linguistique semble donc être précoce. Ce constat remet en cause la séquentialité du modèle tripartite de Morris [28], selon lequel sémantique et pragmatique interviendraient successivement après une analyse syntaxique prééminente et indépendante du contenu sémantique exprimé. Des travaux linguistiques récents tendent de plus à nier l'existence même de niveaux de connaissance linguistique distincts. A titre d'exemple, Rastier a ainsi caractérisé de nombreuses interférences entre fonctions syntaxiques et relations micro-sémantiques [32]. Le système linguistique se présente donc comme une entité intégrée, dans laquelle informations syntaxiques, sémantiques et pragmatiques se recouvrent largement .

D'un point de vue computationnel, il se pose alors la question du choix du système de description des connaissances linguistiques. Une première solution consiste à intégrer connaissance grammaticale et syntaxique dans un même module. Le système DIAL [30] associe ainsi une grammaire locale à chaque cas sémantique. Les grammaires fonctionnelles d'unification intègrent aussi sous un même formalisme connaissance lexicale et structurelle [20]. Si une telle solution reste séduisante, on peut s'interroger sur son adéquation à des domaines applicatifs relativement vastes. En effet, la définition exhaustive des relations entre syntaxe et sémantique est un travail fastidieux qui est contourné le plus souvent par des solutions *ad hoc*. En conséquence, nous avons préféré mettre en œuvre l'analyse linguistique sous la forme d'une forte coopération entre deux agents concurrents responsables respectivement des analyses syntaxique et sémantique. Cette contrainte de

coopération préserve la nature intégrée du processus linguistique et favorise un comportement adaptatif.

5.3. STRATEGIE COOPERATIVE

L'analyse linguistique dans le système MICRO est réalisée en deux étapes. Dans un premier temps, le système effectue une reconnaissance structurale durant laquelle l'élaboration de la structure relationnelle des énoncés contraint la reconnaissance des mots de la phrase. Dans un second temps l'étape d'interprétation fournit une représentation du sens associé au discours. Seule l'étape de reconnaissance structurale a été étudiée à ce jour : elle est mise en œuvre par un analyseur syntaxique et un module d'amorçage sémantique (figure 3). Ces deux agents produisent parallèlement les deux structures relationnelles suivantes :

□ la *structure fonctionnelle* (f-structure) qui décrit les fonctions syntaxiques de surface reliant entre eux les différents constituants syntagmatiques des phrases analysées. La f-structure reprend le formalisme des grammaires lexicales fonctionnelles (ou LFG [5]). Voici par exemple la structure fonctionnelle de la phrase "On déplace la porte gauche" ci-dessous :

```
( DEPLACER
( Suj          ( ON
( COD          ( PORTE
                ( Det          ( LA
                ( Adj           ( GAUCHE
```

□ la *structure sémantique* (s-structure) qui décrit les relations de sens existant entre les mots de la phrase. Ces relations caractérisent les rôles thématiques des mots et présentent de nombreuses ressemblances avec les cas sémantiques [11]. Voici la structure sémantique de la phrase précédente :

```
( DEPLACER
( Agent        ( ON
( Objet        ( PORTE
                ( Attribut     ( GAUCHE
```

A ce niveau, seule la composante relationnelle de la sémantique est utilisée. C'est l'étape d'interprétation qui élabore le sens final de la phrase par inférence sur la structure produite, la sémantique des unités lexicales et le contexte pragmatique. La comparaison des structures produites sur cet exemple très simple dénote l'existence de correspondances entre relations fonctionnelles et sémantiques. [3] et [13] ont montré la réalité psycholinguistique de ces similitudes. Nous avons ainsi défini un KS "Structure Relationnelle" qui fusionne les structures concurrentes suivant un critère de cohérence interne, ainsi que vis à vis des informations contextuelles fournies par les niveaux supérieurs (figure 3). La stratégie coopérative est réalisée par la fusion des productions ascendantes (f-structure et s-structure) et descendantes (prise en compte simultanée de contraintes grammaticales et sémantiques par le KP Accès Lexical) des deux analyseurs. L'analyse syntaxique conserve cependant une priorité en temps normal, en regard de ses capacités structurales, tandis que le KP Sémantique sert de renforçateur d'hypothèses. Ce n'est qu'en cas d'énoncés non normés que ce dernier acquiert toute son importance. On remarque ainsi sur la figure 4 qu'il élabore seul la structure relationnelle des phrases agrammaticales.

Notre ambition à moyen-terme est d'accroître le rôle de l'analyse sémantique aux cas grammaticaux. La prééminence syntaxique serait alors limitée à l'articulation des propositions.

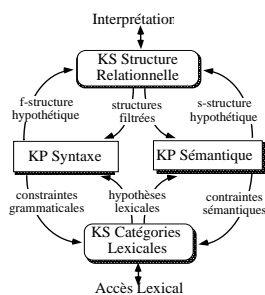


Figure 3 : coopération syntaxe-sémantique

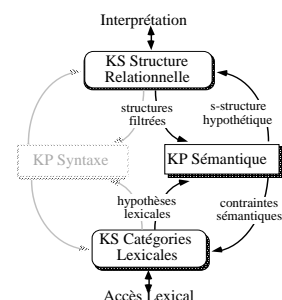


Figure 4: phrases agrammaticales

5.4. KP AMORCAGE SEMANTIQUE

Le KP Syntaxique est dérivé de *Ln_2_3*, un analyseur LFG développé en LISP [38]. Sa mise en oeuvre ne présentant aucune spécificité, nous ne décrivons que l'analyseur sémantique.

5.4.1. Amorçage Sémantique

L'amorçage sémantique peut être schématiquement décrit comme un processus cognitif dans lequel un mot en appelle d'autres par association de sens [3]. Il nous semble cependant que l'on peut distinguer deux types d'amorçage, selon la composante sémantique mise en jeu.

□ *Amorçage relationnel* : il consiste à satisfaire les rôles thématiques des mots de la phrase (décrits dans la structure sémantique). Pour cela, une structure argumentative est associée à chaque lexème³. Elle décrit les relations sémantiques que ce dernier partage avec les autres mots de la phrase, l'amorçage consistant alors à activer les concepts pouvant remplir la structure des lexèmes déjà reconnus. Voici par exemple la structure associée au verbe "déplacer" :

[Agent, Objet, (Destination)]

Ici, la caractérisation d'un agent et d'un objet est obligatoire, alors que la destination est facultative. D'autres rôles non spécifiques au concept peuvent être considérés, mais leur influence sera moindre. Nous avons caractérisé un jeu de 14 cas sémantiques sur un corpus de dialogue regroupant 519 occurrences d'amorçage relationnel. Le système retenu présente de fortes similitudes avec celui de [4]. Une étude de fréquence montre l'existence de 6 relations principales⁴ :

AGENT	"On agrandit cette fenêtre"	f = 0,90
SUJET	"Les fenêtres sont alignées"	f = 0,10
OBJET	"Déplace le curseur"	f = 0,55
ACTIVITE	"On a la possibilité DE sortir"	f = 0,36
ATTRIBUT	"L'endroit désiré"	f = 0,30
MERONYMIE	"La taille DE la fenêtre"	f = 0,30
LIEU / DESTINATION	"Amène le A droite"	f = 0,32
LIEU / SOURCE	"Je veux sortir D'ici"	f = 0,01
MANIERE	"On duplique EN cliquant"	f = 0,10
POSSESSION	"On a le choix"	f = 0,09
INSTRUMENT	"Colorie AVEC le pinceau"	f = 0,06
PATIENT / BENEFICIAIRE	"Cela nous permet de partir"	f = 0,05
BUT	"Agrandit POUR mieux voir"	f = 0,03
TEMPS	"Ensuite, on calcule..."	f = 0,03

□ *Amorçage isotopique* : ce processus permet une focalisation sur des concepts présentant une analogie de sens avec les lexèmes déjà rencontrés. Nous l'utilisons surtout pour rendre compte du phénomène d'isotopie traduisant la redondance thématique du dialogue. Celui-ci se déroule en effet suivant un focus sémantique qui évolue lentement au cours du temps, sauf en cas de rupture manifeste. L'analyse de l'isotopie permet de limiter l'amorçage sur une partie du lexique sémantique correspondant au focus d'attention. Une étude de corpus de dialogue nous a permis de caractériser deux domaines sur lesquels l'utilisateur se focalise alternativement. Ce sont le domaine de la tâche (concepts architecturaux) et le domaine épistémique (concepts liés à l'utilisation de l'ordinateur). Ce processus additionnel d'amorçage isotopique permet de moduler l'amorçage relationnel et réalise une analyse plus riche et plus fine que celle proposée par les grammaires de cas.

5.4.2. Lexique Sémantique Compositionnel

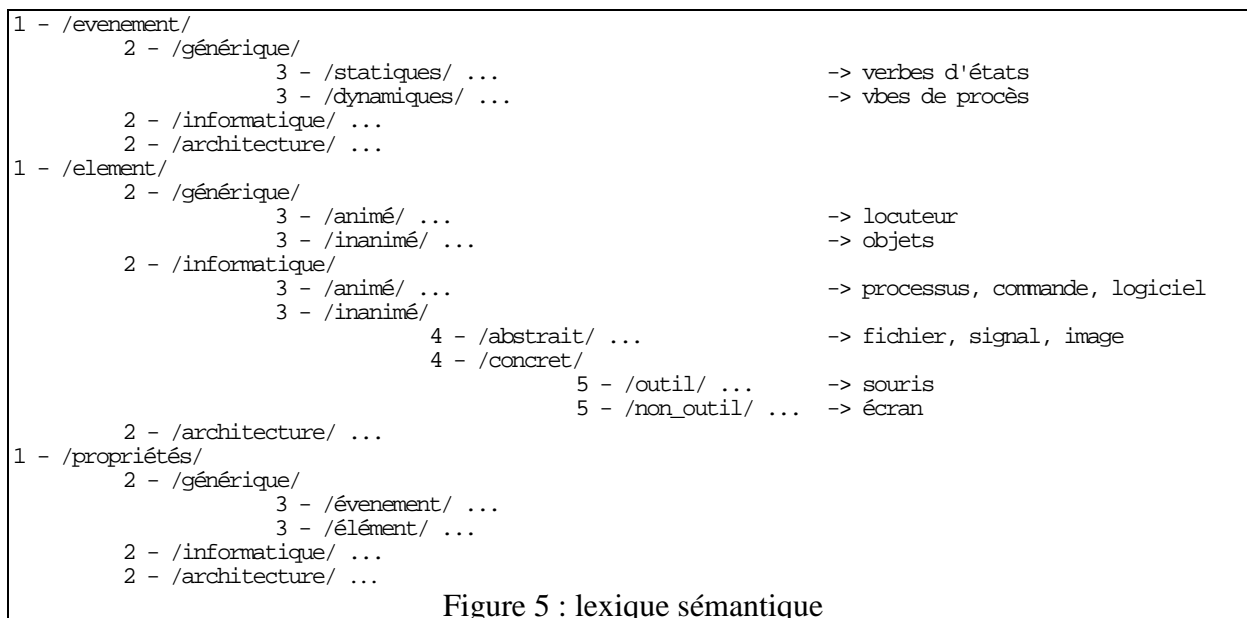
Nous avons représenté l'information sémantique suivant le paradigme compositionnel, selon lequel les unités minimales de sens sont des entités sub-lexicales appelées sèmes [16]. Ces sèmes sont organisés hiérarchiquement suivant leur degré de généralité et sont définis en opposition de sens (exemple : /anime - inanimé/). Le sens lexical correspond alors à un ensemble de sèmes,

³lexème: unité de sens correspondant à une entrée lexicale.

⁴ les mots en relations sont en italiques, tandis que on trouve en majuscules les ligateurs pré-indiquant certains cas. On donne en regard la fréquence d'apparition des cas (nombre d'occurrences moyen par proposition).

souris : /élément/ + /informatique/ + /inanimé/ + /concret/ + /outil/

et le lexique sémantique est organisé sous la forme d'une taxonomie dans laquelle les liens hiérarchiques représentent des relations d'hyponymie/hyperonymie. La figure 5 explicite à titre d'exemple une partie du lexique. Le premier niveau de disjonction correspond aux sèmes macrogénériques qui partagent le champ sémantique en trois *dimensions* : /événement/, /élément/, /propriétés/. Le second niveau est celui des sèmes mésogénériques qui font la distinction entre trois *domaines* : les termes généraux, ceux attachés à la tâche et les concepts informatiques. C'est à ce niveau qu'intervient l'isotopie, caractérisée par une localisation prolongée sur un domaine particulier. Les niveaux suivants regroupent enfin les sèmes plus spécifiques [33].



Le lexique comprend de plus une description des relations sémantiques sur laquelle s'opère l'amorçage relationnel. Un lien "Agent" sera ainsi défini entre le concept de /déplacement_objet/ et les éléments animés du domaines général (le locuteur). Du point de vue de l'amorçage, l'intérêt de cette représentation réside dans sa nature relationnelle. De plus, les sèmes ne sont pas nécessairement des primitives, ce qui garantit une description minimale de la connaissance sémantique⁵.

5.4.3. Description du KP Amorçage Sémantique

Le KP Amorçage sémantique est réalisé sous la forme d'un réseau associatif à six couches (fig. 5.4). Les activations en entrée représentent la force d'amorçage de chaque lexème. Le dernier mot analysé a une activation maximale tandis qu'un mécanisme d'oubli temporel réduit les activations de l'ensemble des lexèmes au cours de l'analyse. L'activation est alors propagée jusqu'à la couche de sortie, où les cellules d'activités maximales correspondent aux mots amorcés. Ce réseau est obtenu par compilation du lexique-sémantique, et non par un mécanisme d'apprentissage. Chaque couche cachée de l'analyseur a ainsi un rôle bien déterminé, à la différence des systèmes neuromimétiques.

⁵ Notre lexique comprend ainsi 154 lexèmes décrit à l'aide d'une trentaine de couple de sèmes disjonctifs.

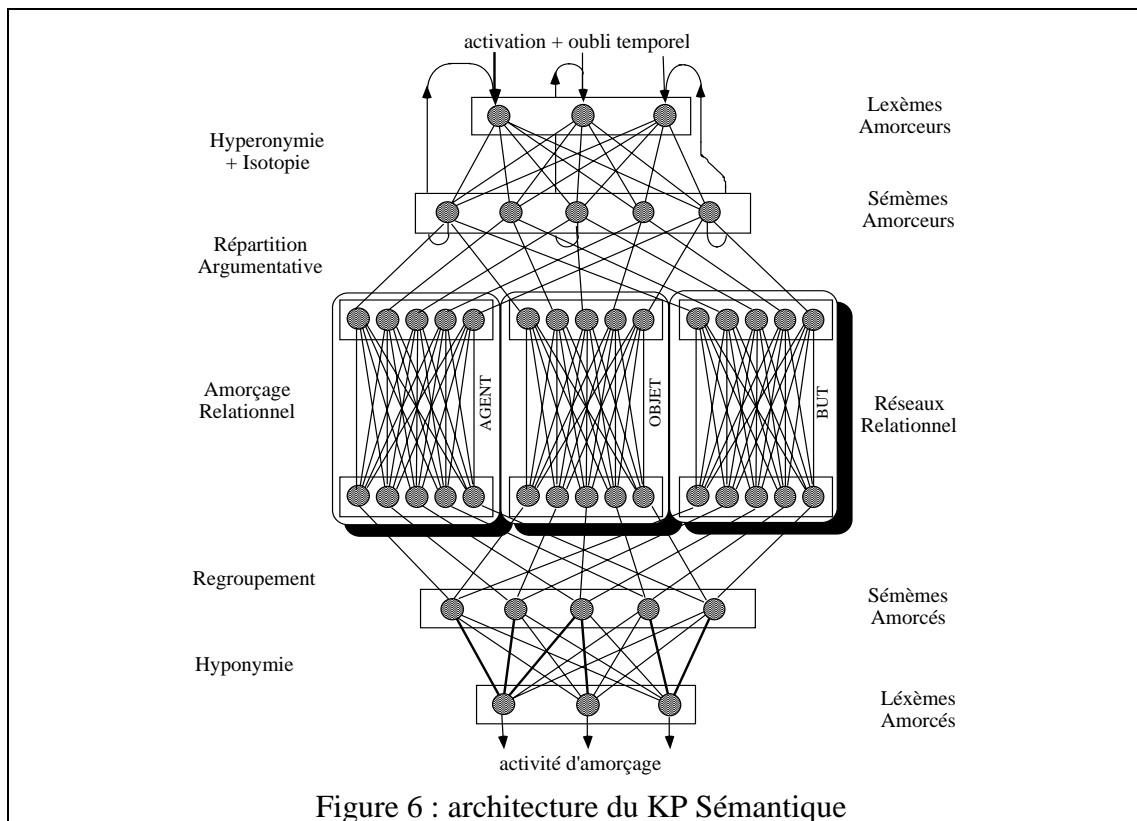


Figure 6 : architecture du KP Sémantique

L'amorçage se déroule comme suit :

- *Excitation hyperonymique et inhibition isotopique* : les activations de la couche d'entrée sont transmises vers la couche suivante représentant les sémèmes⁶ du lexique. L'activation est ainsi propagée, suivant les liens hyperonymiques, des lexèmes vers leurs ascendants hiérarchiques dans la taxonomie. De plus, des liens d'inhibition latérale ont été ajoutés entre les sémèmes de différents domaines. Ils modélisent l'amorçage isotopique entre concepts du domaine courant, au détriment des autres.
- *Répartition argumentative* : cette étape prépare l'amorçage relationnel. Les activations sémémiques sont réparties, à partir des structures argumentatives, entre les différentes relations sémantiques. Dans l'exemple du verbe "déplacer", les liens obligatoires "Agent" et "Objet" recevront ainsi la totalité de l'activation sémémique, à la différence des liens facultatifs. Les activations obtenues nourrissent des réseaux internes propres à chaque relation (figure 6). On opère ainsi une différenciation des liens d'amorçage nécessaire à l'élaboration de la structure sémantique.
- *Ligateurs* : A ce stade interviennent les ligateurs, qui jouent un rôle de focalisateurs sur des relations particulières. La préposition "PAR" sert par exemple à pré-amorcer différents rôles sémantiques:

"Entre PAR le menu d'accueil"	: localisation
"On écoute PAR la touche F5"	: instrument
"Il est affiché PAR le logiciel"	: agent

Des cellules (non représentées) correspondant à chaque ligateur, interviennent alors comme entrées d'adaptation contextuelle permettant de moduler l'activité des réseaux relationnels internes.

- *Amorçage* : cette phase correspond à l'amorçage relationnel proprement dit. Les activations sémémiques d'entrée sont propagées aux sémèmes de sortie partageant avec eux le lien considéré.
- *Regroupement des amorces* : c'est l'opération inverse de la répartition argumentative. Chaque sémème de l'avant dernière couche reçoit les activités sémémiques de chaque réseau relationnel

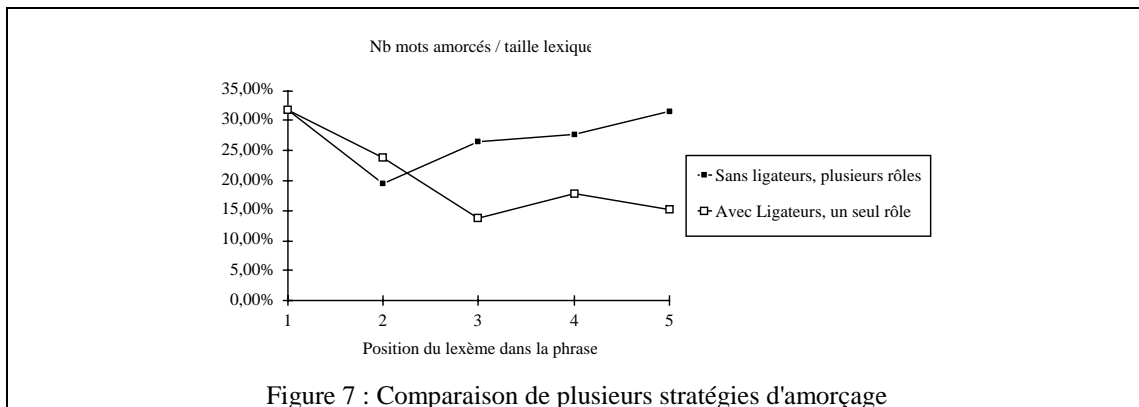
⁶ un sémème est une entité conceptuelle représentée par un noeud de la taxonomie. Les sèmes n'étant pas primitifs, les sémèmes sont dépendants de la description choisie. On peut considérer les lexèmes comme des sémèmes nominalisables.

interne. L'activité résultante est obtenue par une heuristique de type Min/Max afin de favoriser une forte activité sur un lien. On conserve plusieurs liens saillants car un mot peut jouer plusieurs rôles sémantiques.

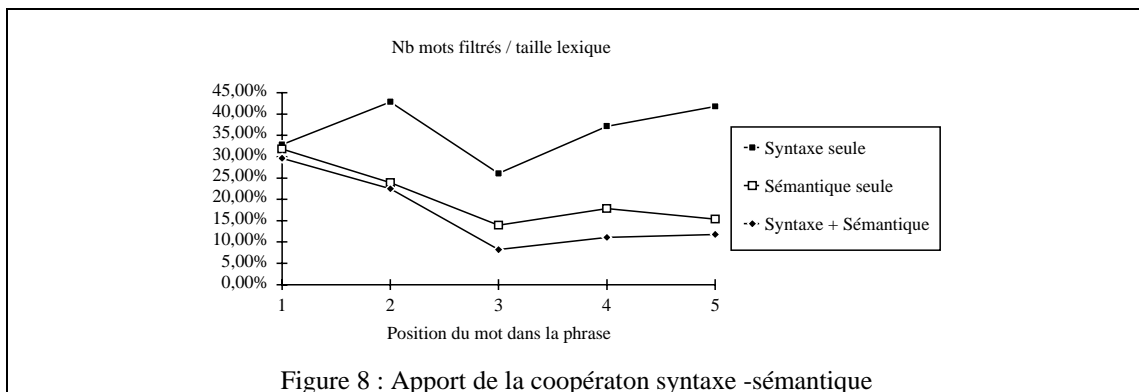
□ *Transfert hyponymique* : ici, l'activité des sémèmes est reportée par les liens hyponymiques vers les lexèmes de la couche de sortie. Le KP Sémantique conserve plusieurs lexèmes parmi les plus saillants, afin de permettre une recherche en largeur d'abord. Ce sont les mots amorcés, qui seront insérés dans la structure sémantique hypothétique, suivant leur(s) relation(s) d'amorçage(s).

5.4.3. Résultats

Nous avons effectué plusieurs analyses en raffinant à chaque fois le mécanisme d'amorçage (ajout des ligateurs et limitation du nombre de rôles que peut satisfaire un lexème). La figure 7 illustre le taux de filtrage lexical obtenu pour chaque méthode suivant l'axe syntagmatique. On voit ainsi l'apport des ligateurs dès le troisième lexème de la phrase (les relations précédentes ne donnent généralement pas lieu à un pré-amorçage par ligateur) : on arrive alors à filtrer 85 % du lexique.

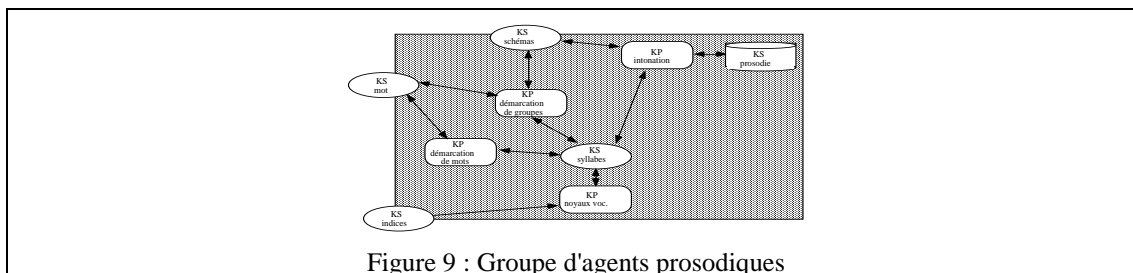


On note ainsi que ce filtrage sémantique est sensiblement meilleur que celui qui est obtenu par une analyse syntaxique seulement limitée au filtrage catégoriel (Nom, Verbe..), c'est à dire portant uniquement sur les radicaux morphologiques. La figure 8 montre l'apport de la coopération entre les deux analyses.



6. VOIE HOLISTIQUE

La voie holistique du système repose sur une analyse prosodique (figure 9) qui a pour objectif de fournir rapidement des informations qui doivent permettre de synchroniser les traitements du groupe linguistique sur des îlots de confiance (KP démarcation de mots et de groupe).



Parallèlement, il fournit des informations sur les actes de dialogue (KP intonation). Ces informations sont déduites des noyaux vocaliques (en général centres des voyelles) calculés sur le signal.

6.1. KP Noyaux vocaliques.

6.1.1. Description

D'une part, cet agent détecte les noyaux vocaliques à partir des courbes d'énergie et de fréquence fondamentale fournies par les agents perceptifs. D'autre part il met à la disposition des agents de plus haut niveau un ensemble de paramètres utiles pour le repérage des groupes prosodiques :

- l'énergie moyenne sur chaque noyau,
- la fréquence fondamentale moyenne,
- la durée de chaque noyau,
- l'accélération du débit d'un noyau au suivant,
- la durée de pause éventuelle précédant le noyau.

La détection des noyaux vocaliques est un traitement-clef puisque les autres agents s'appuient sur cette information. L'analyse holistique se doit d'être rapide, afin de servir de pré-marquage à la voie analytique en anticipant les points d'articulation de la prise de décision. Afin de satisfaire ces contraintes de rapidité, la détection des noyaux vocaliques se fait uniquement par une recherche de formes spécifiques (pics et plateaux) sur la courbe d'énergie. L'étude présentée au paragraphe suivant montre que ces contraintes ne nuisent pas à la robustesse.

6.1.2. Robustesse de la détection

Deux analyses ont été effectuées. La première a été réalisée sur un corpus lu extrait de BDSON [34] et montre une robustesse très satisfaisante. De plus, un grand nombre de noyaux doubles (voyelles géminées dans le même noyau) ne pénalisent pas les traitements ultérieurs.

Nombre de noyaux non détectés	Nombre de noyaux détectés mais faux	Nombre de noyaux détectés mais doubles	Nombre total de noyaux
11	25	58	944
1.1 %	2.6 %	6.1 %	

La seconde analyse a été effectuée sur un corpus de dialogue enregistré dans un environnement bruyé. Là encore, les résultats sont satisfaisants. L'agent responsable de la détection des noyaux vocaliques présente ainsi une bonne robustesse d'ensemble.

Nombre de noyaux non détectés	Nombre de noyaux détectés mais faux	Nombre de noyaux détectés mais doubles	Nombre total de noyaux
3	8	10	192
1,6 %	4,2 %	5.2 %	

6.2. KP DEMARCATION DE MOTS

Cet agent est chargé de placer des marqueurs en début et fin de mots ; s'il le peut, il doit aussi qualifier les mots marqués en termes de *mot lexical* ou de *mot grammatical*. Cette tâche est réalisée

par l'application d'un ensemble de règles prosodiques issues des travaux réalisés dans ce domaine par [31] à la suite de l'expertise fournie par [9]. Voici un exemple de règle :

```

SI Fo (n) = MAX_Fo
  ET D(n-2) • D(n-1) < D(n)
  ET D(n) • (D(n-1) x 1.5)
  ET D(n) = 4
ALORS FML(n)

```

Fo(n) = fréquence fondamentale quantifiée en niveaux sur le n° noyaux vocalique.

D(n) = durée quantifiée du n° noyau vocalique.

FML(n) = marque de Fin de Mot Lexical posée sur le n° noyau.

Nous travaillons sur un affinement de ces règles pour les adapter à un environnement de dialogue.

6.3. KP DEMARCATIION DE GROUPES

Cet agent doit déterminer les début et fin de groupe syntaxique à partir des mots démarqués précédemment et des informations fournies par le *KP noyaux vocaliques*. Il doit aussi déterminer le type des groupes repérés : groupe majeur ou mineur dans la phrase. C'est un agent à base de règles qui a fait l'objet d'une étude analogue à la précédente.

7. ETUDE DU CONTROLE

7.1. PROBLEMATIQUE

Dans une architecture multi-agents, la maîtrise des processus de contrôle est de la première importance pour un ordonnancement correct des tâches des agents. Afin de caractériser ce contrôle, nous avons choisi de simuler le travail d'un groupe d'agents par une série d'expertises humaines : nous l'avons fait pour le groupe acoustico-phonétique.

Deux rôles ont été dévolus à l'expert : (a) il simule le fonctionnement des agents à l'aide d'interfaces graphiques, afin de garantir la qualité des traitements (b) il fixe la stratégie d'unification des hypothèses ascendantes et descendantes. Cette unification, à la base de l'interactivité, conditionne la prise de décision globale et reste donc fondamentale dans le système. Les agents Acoustico-Phonétique et Microprosodie travaillent sur des domaines de compétence disjoints. Par contre, les hypothèses phonémiques produites par le KP Phonologie englobent ces deux domaines (voir tableau). Le processus d'unification doit donc posséder des mécanismes d'union et d'intersection d'hypothèses très fiables.

Hypothèses ascendantes		Hypothèses descendantes	
KP Acoustico-Phonétique	Macro-classe	Macro-classe	KP Phonologie
	Traits 1 ... 4	Traits 1 ... 4	
KP Micro Prosodie	Voisement	Voisement	
	Energie	Energie	
	Durée intrinsèque	Durée intrinsèque	
	Traits calculés sur le signal	Traits typiques aux phonèmes	

7.2. INTERFACES DE SIMULATION

L'expert travaille sur des interfaces représentant l'état de chaque agent KP simulé (figures 9 et 10).

7.3. RESULTATS D'EXPERTISE

Dans un premier temps, l'expert a travaillé sans contraintes. Nous voulions dégager une expertise sur la stratégie d'unification la plus naturelle possible. Rien n'assure cependant que la meilleure stratégie pour un système automatique corresponde à celle qui est employée par l'expert humain. C'est pourquoi nous avons ensuite imposé à l'expert diverses stratégies d'unification pour comparer leurs performances. L'expertise a porté sur un corpus extrait de BDSONS [34].

7.3.1. Expertise libre

Utilisation et robustesse des traits : pour chaque trait, nous avons comptabilisé le nombre de fois où l'expert a pu (ou a estimé nécessaire de) qualifier ce dernier par un attribut :

Nom du trait	Macro-classe	Trait 1	Trait 1	Trait 1	Trait 1	Voisement	Energie
Utilisation	100 %	75,4 %	36,8 %	3,5 %	0,0 %	96,5 %	86,0 %

Les traits "Macro-classe", "Voisement" et "Energie", robustes, sont ainsi très utilisés par l'expert. Ces deux derniers n'apportent souvent qu'une confirmation à l'information portée par "Macro-classe" (forte redondance). Enfin, les traits 3 et 4 ne sont pas utilisés, du fait de leur manque de robustesse.

Confiance accordée aux traits : nous avons considéré la confiance que l'expert accorde à chaque trait au moment de l'unification. Pour cela, nous avons calculé trois taux de confiance :

- *Confiance forte (OK+)* : l'expert a conservé uniquement les hypothèses descendantes compatibles avec le trait considéré.
- *Confiance moyenne (OK-)* : en sus des hypothèses précédentes, l'expert a conservé celles qui ne contredisaient pas ce trait. Un taux OK Total regroupe OK+ et OK-
- *Confiance nulle (NO)* : l'expert a conservé au moins une hypothèse entrant en contradiction avec le trait considéré.

Nom du trait	OK+	OK-	OK Tot	NO
Macro-classe	78,9	3,5	82,4	17,6
Trait 1	26,8	22,0	48,8	51,2
Trait 2	23,8	23,8	47,6	52,4
Voisement	92,7	-	92,7	7,3
Energie	83,7	-	83,7	16,3

On remarque que l'expert accorde une grande confiance aux traits "Macro-classe", "Energie" et "Voisement". Si de plus, aucun candidat du KP Phonologie n'est compatible avec ces traits, toutes les hypothèses sont conservées. Dorénavant ces traits sont appelés *traits principaux*. En revanche, l'expert n'accorde qu'une confiance limitée aux autres traits appelés dans la suite *traits secondaires*.

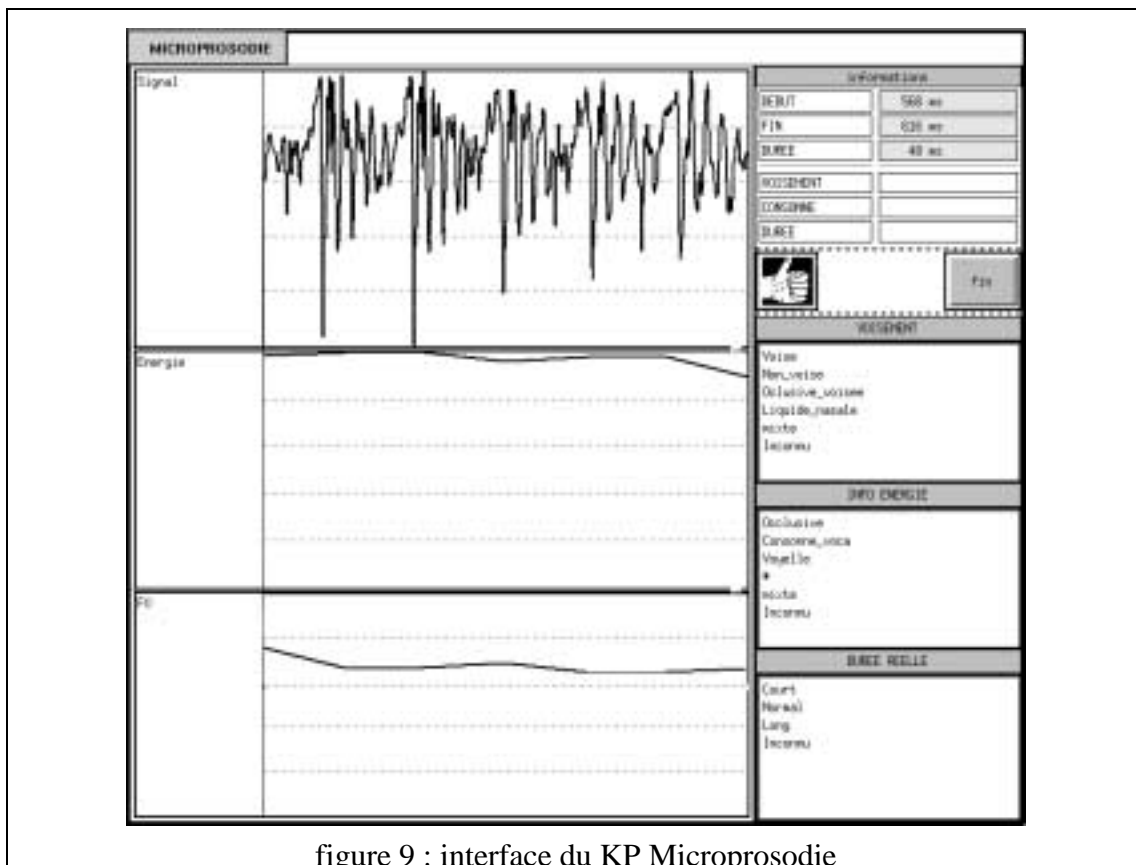


figure 9 : interface du KP Microprosodie

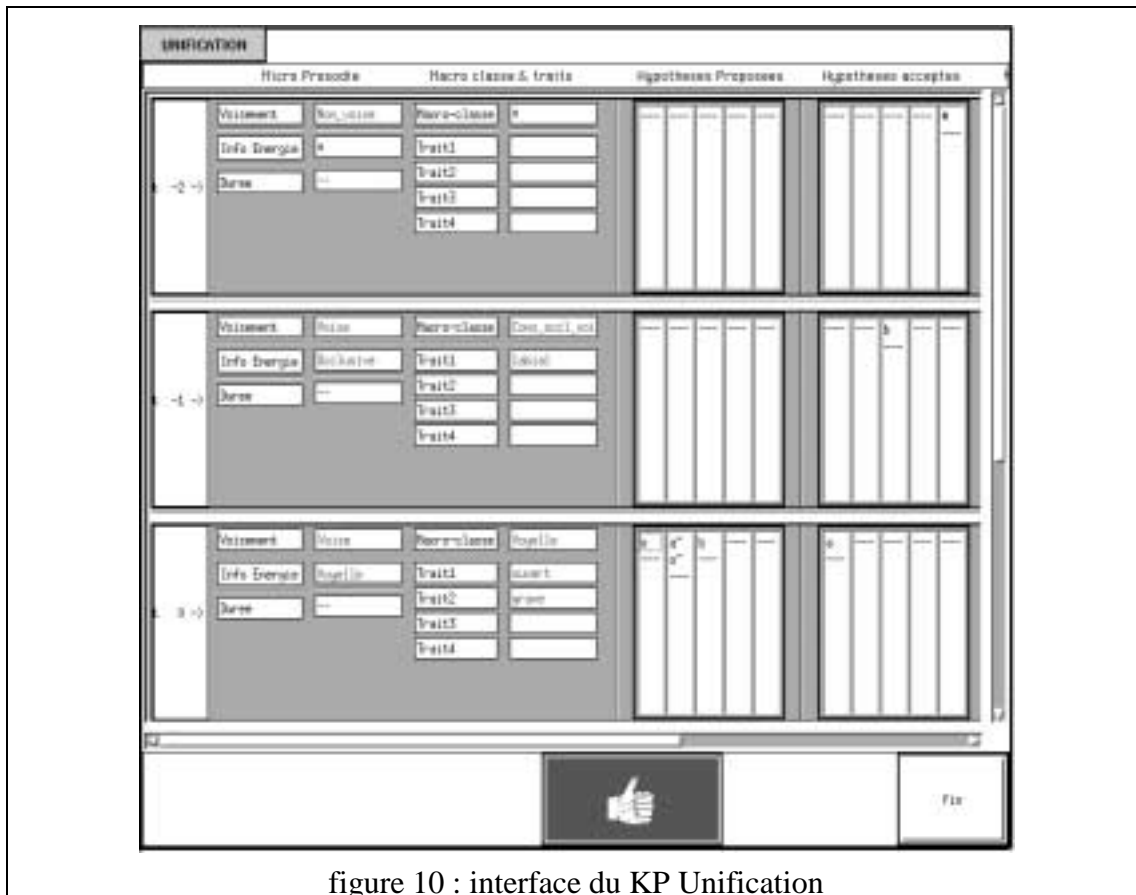


figure 10 : interface du KP Unification

7.3.2. Stratégie d'unification de l'expert

L'étude précédente montre que l'expert préfère adopter une stratégie prudente qui consiste à garder un maximum d'hypothèses, en éliminant progressivement, à l'aide des traits principaux, les hypothèses erronées. Dans de rares cas contextuellement caractérisés, l'expert a choisi une stratégie différente.

7.3.3. Expertise contrainte

A partir des comportements de l'expert, nous avons pu étudier trois stratégies particulières :

- *stratégie normale* : c'est la plus proche de la stratégie d'expertise libre. L'expert garde une hypothèse compatible avec tous les traits principaux et deux traits secondaires,
- *stratégie large* : l'expert garde une hypothèse compatible avec tous les traits principaux,
- *stratégie stricte* : l'expert garde une hypothèse compatible avec tous les traits.

7.3.4. Analyse qualitative des stratégies : nous avons caractérisé la qualité⁷ des stratégies à l'aide d'un critère pondérant la robustesse des résultats par le nombre d'hypothèses finales produites :

Stratégie	Stricte	Normale	Large
Robustesse pondérée	54,2 %	64,2 %	63,1 %

Les écarts de performances se sont réduits, montrant qu'on ne peut préférer une stratégie à une autre dans toutes les situations. En outre, leurs performances globales restent insuffisantes. D'où la nécessité de définir une stratégie évolutive s'adaptant au contexte.

7.4. POUR UNE STRATEGIE ADAPTATIVE

⁷ Dans le cas de la simulation, la qualité des hypothèses lexicales reflète totalement celle du groupe acoustico-phonétique, car le groupe linguistique ne fournit ici aucune contrainte sur celles-ci.

Nous nous sommes ainsi orienté vers une stratégie adaptative, mélangeant selon le contexte les trois stratégies précédentes. Nous avons alors déterminé quatre ensembles de paramètres pour l'adaptation contextuelle de la stratégie :

- les paramètres d'environnement acoustique,
- la concordance informative entre les traits,
- le type de phonotype candidat,
- la trace du comportement des agents KP.

En jouant sur ces paramètres il est possible d'obtenir des critères pour une méta-stratégie pilotant la stratégie de contrôle. Notre étude porte désormais sur la caractérisation de l'adaptation à ces informations.

8. CONCLUSION

Nous avons donné dans cet article un aperçu des études faites dans le cadre du GDR-PRC "Communication Homme-Machine" à propos du système de compréhension automatique MICRO. Ce système adopte une approche fondée sur la cognition en y distinguant des principes de modularité, d'interactivité et de double analyse. Les nombreux modules, spécialisés et indépendants, échangent leurs informations par diffusion. Les conflits dans les échanges nécessitent des régulations. La prise de décision délocalisée nécessite de son côté de connaître les effets produits à plus long terme dans les étages éloignés du système. Pour ces deux raisons le contrôle joue un rôle important dans le système : il a pour charge de réduire l'ambiguïté tout en permettant au système de s'adapter c'est-à-dire de prendre en compte des situations nouvelles ou inattendues. L'étude présentée ici a conduit à des acquis importants pour cette architecture bien qu'une expérimentation systématique n'ait pas encore été faite.

BIBLIOGRAPHIE

- [1] J.Y. Antoine, B. Caillaud, J. Caelen, 1993, Syntax-semantics cooperation in MICRO, a multi-agents speech understanding system, EUROSPEECH'93, Berlin
- [2] O. Baujard et C. Garbay, 1990, A programming environment for distributed expert system desing, Expersys, 27-32.
- [3] T.G. Bever, 1970, The cognitive basis for linguistic structures, J.R. Hayes (ed.), Cognition and the development of language, New-York, 279-362
- [4] B. Boguraev B, K. Spark-Jones, 1987, A Note on a Study of Cases, Computational Linguistics, Vol 13, n° 1-2, pp. 65-68.
- [5] J. Bresan et R. Kaplan, 1981, Lexical Functional Grammars; a Formal System for Grammatical Representation, Halle, Bressan & Miller, MIT Press, Cambridge, Mass.
- [6] M.P. Bryden, 1988, Cerebral specialization: clinical and experiment assement, dans F. Boller et J. Grafman (Eds), Handbook of Neuropsychology, 1, Elsevier Publ., 143-159.
- [7] J. Caelen, 1979, "Un modèle d'oreille. Analyse de la parole. Reconnaissance phonémique, Thèse d'Etat, Toulouse.
- [8] J. Caelen et al., 1990, Architecture et fonctionnement du système DIRA. De l'acoustique aux niveaux linguistiques", Revue de traitement du signal, vol. 7, no 4, pp. 345-366.
- [9] G. Caelen-Haumont, 93, "Cognitive processes and prosodic encoding. Speaker's adaptation to discourse conditions", Communication and cognition - Artificial Intelligence, 10:4.
- [10] B. Caillaud, J.Y. Antoine et J. Caelen, 1994, MICRO, un système multi-agent pour la compréhension de la parole, actes 9° RFIA, Paris.
- [11] C. Fillmore, 1968, The Case for Case, in Universal in Linguistic Theory, Bach & Harms, Chicago, 1-90.
- [12] J.A. Fodor, 1983, The modularity of Mind, MIT Press, Cambridge, Mass.
- [13] K.I. Forster, 1979, Levels of Processing and the Structure of the Language Processor, in Cooper W.E. and Walker E.C.T. (Eds), Sentence Procesing: Psycholinguistic Studies, 216-225.
- [14] W.R. Gardner, 1974, The processing of information and structure, Potomac, Md, Erlbaum.
- [15] R.W. Gardner, 1981, The analysis if unanalysed perceptions, dans M. Kubovy, J.R. Pomerantz (Eds), Perception and organization, Lawrence Erlbaum Ass., Hillsdale, N.J., 119-139.
- [16] A. Greimas, 1966, Sémantique structurale, Larousse, Paris.
- [17] B. Hayes-Roth, 1985, Blackboard architecture for control, Journal of Artificial Intelligence, 26, 251-321.

- [18] M. Imbert, 1992, Neurosciences et sciences cognitives, in D.Andler (dir.) Introduction aux sciences cognitives, Gallimard, Paris, 49-76.
- [19] J.F. Jodouin, 1993, Réseaux de neurones et Traitement du langage naturel: étude des réseaux de neurones récurrents et de leurs représentations, Thèse Paris XI Orsay.
- [20] M. Kay, 1979, Functional grammars, 5th meeting of the Berkeley linguistic society, 142-158.
- [21] Kosslyn, 1990, Cognition, 34, 203-277.
- [22] J. Lautrey, 1990, Esquisse d'un modèle pluraliste du développement cognitif, M. Reuchlin et al (Eds.), Cognition: l'individuel et l'universel, PUF, Paris, 185-216.
- [23] D. Marr, 1982, Vision, W.H. Freeman, San Francisco, California.
- [24] J.L. McClelland, J.L. Elman, 1986a, The TRACE model of speech perception, Cognition, 18, 1-86.
- [25] J.L. McClelland, D.E. Rumelhart and the PDP Group, 1986b, *Parallel Distributed Processing: explorations in the microstructure of cognition*, MIT Press/ Bradford books, Cambridge, Ma.
- [26] W.D. Marslen-Wilson, A. Welsh, 1982, Processing interactions and lexical access during word recognition in continuous speech, Cognition, 10, 29-63.
- [27] M. Minsky, 1985, *The society of Mind*, Simon & Schuster, New-York.
- [28] C. Morris, 1946, *Signs, Language and Behavior*, Prentice Hall, NY
- [29] J. Morton, 1983, *Le lexique interne*, La Recherche, 14, 143, 474-481.
- [30] P. Moussel, J.M. Pierrel, A. Roussalany, 1989, *Cooperation and Representation of Syntactic-Semantic and Pragmatic Knowledge in a natural Task Oriented Spoken Dialogue System*, EUROSPEECH'89, Paris.
- [31] M.K. Nasri, G. Caelen-Haumont et J. Caelen, 1989, "Using Prosodic Rules in Continuous Speech Recognition", Proc. of ICASSP-IEEE, Glasgow, Vol. 1, pp. 671-674.
- [32] F. Rastier, 1988, *Microsémantique et Syntaxe*, L'information Grammaticale, 37, p. 8-13.
- [33] F. Rastier, 1987, *Sémantique Interprétative*, PUF, Paris
- [34] J.F. Sérignat, J.M. Dolmazon, M. De Calmes, G. Pérennou, 1992, "BDSONS - BDLEX : Bases de données vocales et lexicales", Rencontres Recherches en Informatique - Industrie, Toulouse, PP 88-89.
- [35] R.W. Sperry, 1961, *Cerebral Organization and behavior*, Science, 133, 1749-1757.
- [36] A. Treisman, 1992, *L'attention, les traits et la perception des objets*, in D.Andler (dir.) *Introduction aux sciences*
- [37] D.W. Zaidel, 1984, Les fonctions de l'hémisphère droit, La Recherche, 15 (153), 332-340.
- [38] P. Zweigenbaum, M. Cavazza, 1990, *Deep-text understanding in a Restricted Domain*, COLING, Helsinki.
- [39] E.B. Zurif, 1974, Auditory lateralization: prosodic and syntactic factors, Brain and Language, 1, 391-404.